

ASSISTANT COMMISSIONER FOR PATENTS
BOX PATENT APPLICATION
Washington, D.C. 20231

09/01/00
JC675 U.S. PTO

09-05-00
A
DOCKET NO. YOR920000607US1
Date: September 1, 2000

Sir:

Transmitted herewith for filing is the Patent Application/of: under Rule 1.53(f)

Inventors: Thomas Anthony Cofino, Stephen H. Gomory, Juhnnyoung Lee,
Mark Edward Podlaseck

For: BUSINESS METHOD FOR VISUALLY ANALYZING CLICKSTREAM DATA
WITH A PARALLEL COORDINATE SYSTEM

Enclosed are:

- ☒ 13 Sheets of Informal Drawings.
- ☐ An assignment of the invention to International Business Machines Corporation, Armonk, New York 10504.
- ☐ A certified copy of a _____ application.
- ☒ Unsigned Declaration and Power of Attorney is attached to the application.
- ☐ Associate Power of Attorney.
- ☒ Information Disclosure Statement with form PTO-1449 with references attached.

The filing fee has been calculated as shown below:

	(Col. 1)	(Col. 2)
FOR:	NO. FILED	NO. EXTRA
BASIC FEE		
TOTAL CLAIMS	21 - 20 =	1
INDEP CLAIMS	1 - 3 =	0
____ MULTIPLE DEPENDENT CLAIM PRESENTED		

OTHER THAN A SMALL ENTITY	
RATE	FEE
	\$ 690.00
X \$ 18 =	\$ 18.00
X \$ 78 =	\$ 0.00
+ \$ 260 =	\$ 0.00
TOTAL	\$ 708.00

If the difference in Col. 1 is less than zero, enter "0" in Col. 2.

- ☒ Please charge my Deposit Account No. 09-0468 in the amount of \$ 708.00.
- ☒ The Commissioner is hereby authorized to charge payment of the following fees associated with this communication or credit any overpayment to Deposit Account No. 09-0468. A duplicate copy of this sheet is enclosed.
- ☒ Any additional filing fees required under 37 CFR 1.16.
- ☒ Any patent application processing fees under 35 CFR 1.17.

Respectfully submitted,

By Louis J. Percello
Louis J. Percello
Registration No.: 33,206
Tel. (914) 945-3145

IBM CORPORATION
INTELLECTUAL PROPERTY LAW DEPT.
P.O. BOX 218
YORKTOWN HEIGHTS, NY 10598

Express Label EL600330839US
Date of Deposit: Sept. 1, 2000

BUSINESS METHOD FOR VISUALLY ANALYZING CLICKSTREAM DATA
WITH A PARALLEL COORDINATE SYSTEM

5

FIELD OF THE INVENTION

This invention relates to shopping on a computer network. More specifically, the invention relates to the visualization and analysis of clickstream data from one or more online stores where clickstream data is visualized and analyzed by using an extended parallel coordinate system.

BACKGROUND OF THE INVENTION

Commerce over networks, particularly e-commerce over the Internet, has increased significantly over the past few years. Part of e-commerce enables users/customers to access information of products and to purchase them from various commercial Web sites (i.e. online stores). There are numerous online stores currently operating in the Internet including: Amazon.com, eToys.com, Buy.com, Wal-Mart.com, LLBean.com, and Macys.com. These online stores provide various customer services to make commerce activities possible over Web sites. Some of the examples of the basic services are catalogs of merchandise which are both browsable and searchable by various

product attributes (e.g., keyword, name, manufacturer, and model number), shopping carts, and checkout process. Some online stores also provide advanced customer services such as wish lists, gift registries, calendars, custom-configuration of products, buyer's groups, chatting, e-mail notification, product evaluations, product recommendations and in-context sales.

5

As shopping experience in the Internet gets deeper and broader, it becomes an important task for merchants of online stores to understand and analyze the shopping behavior of customers and to improve the shopping experience in their online stores by using this analysis. A basic unit for such analysis is clickstream data from online stores. Clickstream is a generic term to describe visitors' paths through one or more Web sites. A series of Web pages requested by a visitor in a single visit is referred to as a session. Clickstream data in an online store is a collection of sessions on the site. Clickstream data can be derived from raw page requests (referred to as hits) and their associated information (such as timestamp, IP address, URL, status, number of transferred bytes, referrer, user agent, and, sometimes, cookie data) recorded in Web server log files. Analysis of clickstreams shows how a Web site is navigated and used by its visitors.

In an e-commerce environment, clickstreams in online stores provide information essential to understanding the effectiveness of marketing and merchandising efforts, such as how customers find the store, what products they see, and what products they buy. (While not all this information may be available from Web server log files, it can be extracted from associated data sources such as commerce server databases and tied together with HTTP request data.) Analyzing such information embedded in clickstream data is critical to improve the effectiveness of Web marketing and merchandising in online stores. Interest in interpreting Web usage data in Web

server log files has spawned an active market for Web log analysis tools that analyze, summarize, and visualize Web usage patterns.

5 PROBLEMS WITH THE PRIOR ART

Interest in interpreting Web usage data in Web server log files has spawned an active market for Web log analysis tools that analyze, summarize, and visualize Web usage patterns. While useful to some extent, most of existing tools have the following shortcomings: (1) the summaries they provide obscure useful detail information, (2) the static displays such as histograms and pie charts restrict users to passive interpretation, and (3) the weak (or lack of) connection between purchase data and navigation data limits the ability to understand the site's effectiveness in terms of return on investment.

10
15
20 OBJECTS OF THE INVENTION

An object of this invention is an improved system and method for showing steps in a process relative to an independent variable.

20 An object of this invention is an improved system and method for showing sequential steps in a process, specifically the clickstream data of a web site.

An object of this invention is an improved system and method for showing click stream data of a web site that does not obscure detailed information.

5 An object of this invention is an improved system and method for showing click stream data on a web site that includes both purchase data and navigation data.

SUMMARY OF THE INVENTION

10 This invention presents an interactive parallel coordinate system that can be used to provide users with greater abilities to interpret and explore clickstream data of online stores on the Web. The interactive parallel coordinate system in this invention uses the multiple axes of the system to represent steps of a process with respect to an independent variable. Specifically, the process steps are sequential steps which sessions take during their navigation of a Web site.

15 In a preferred embodiment, the sessions are displayed as respective polygonal lines on a parallel coordinate system. That is, the clickstream of a session is represented by a polygonal line that intersects its value in each axis selected in the parallel coordinate system. With one or more variables such as referrers and host names that categorize sessions assigned to axes, the system can visualize the relationship between session categories and navigation paths. In one embodiment, by looking at the last axis a polygonal line intersects, it is straightforward to see at what point sessions leave the site. In order to help users actively explore and interpret data of

20

interest, this system provides facilities for filtering, sampling, clustering, color coding, and dynamic querying.

5 **BRIEF DESCRIPTION OF THE FIGURES**

The foregoing and other objects, aspects, and advantages will be better understood from the following non-limiting detailed description of preferred embodiments of the invention with reference to the drawings that include the following:

Figure 1 is a block diagram of one preferred system architecture.

Figure 2 is a flow chart of a preferred business process.

Figure 3 is an example of an online store structure.

Figure 4 is an example of a Web server log of an online store.

Figure 5 is an example of sessions in an online store used in one preferred embodiment.

Figure 6 is an example of micro-conversions in an online store.

Figure 6A is a flow chart of a shopping step finder process.

Figure 7 is an example of a prior art parallel coordinate system.

Figure 8 is an example of micro-conversion visualization with a parallel coordinate system.

5

Figure 9 is an example of micro-conversion visualization with filters.

Figure 10 is an example of a session browsing more than one online stores.

Figure 11 is an example of micro-conversions of a session browsing more than one online stores.

Figure 12 is an example of visualizing sessions visiting more than one online stores.

DETAILED DESCRIPTION OF THE INVENTION

This invention presents an interactive parallel coordinate system that can be used to provide users with greater abilities to interpret and explore clickstream data of online stores on the Web. Parallel coordinates are a visualization method for displaying multivariate data sets to identify the relationship among the variables. A parallel coordinate system comprises a series of parallel lines that are placed perpendicular to the x-axis (or y-axis) of a Cartesian coordinate system, and preferably each adjacent axis is equidistant, although equidistance is not required. Each parallel axis is assigned a specific dependent variable and dependent variable values are plotted along the

respective axis. The independent variable is represented by polygonal lines which connect the corresponding dependent variable values relating to the independent variable and illustrating a relationship between an independent variable and the dependent variables appearing on each axis.

5 The interactive parallel coordinate system in this invention uses the multiple axes of the system to represent sequential steps which sessions take during their navigation of a Web site, and displays a large number of individual sessions as polygonal lines. The session is a set of events (that may be divided into subsets of events. Each of the axes represents one of the events in the session. An independent variable associated with the event represented on the respective axis is plotted for one or more of the events in the session. For example, the clickstream of a session is represented by a polygonal line that intersects its time stamp value for an event represented by the respective axis for the event.

10 In an alternative embodiment, sessions can be further classified. For example, one or more axes of the parallel coordinate system can be used to classify the sessions by having the classifier as a variable of the axis. For example, if the variable of the "classifier axis" is service provider, all sessions initiated from service provider "A" would start from point "A" on the classifier axis while all sessions initiated from service provider "B" would start from point "B" on the classifier axis. For instance, with one or more variables such as referrers and host names that categorize sessions
15
20 assigned to axes, the system can visualize the relationship between session categories and navigation paths.

In one embodiment, the sessions do not always have events plotted on all of the axes. By arranging the axes in time sequence order, a user can determine at which point in the time sequence order each of the sessions ended. If a particular axis is chosen as a "terminal step", e.g. product purchase, this invention gives the user a clear visual indication of how many of the sessions ended at the "terminal step", e.g. how many customer sessions ended in a product purchase. Further, by looking at the axis at which a polygonal line ends, it is straightforward to see at what point sessions leave the site.

In other embodiments, the invention helps users actively explore and interpret data of interest, by providing facilities for filtering, color coding, sub-categorizing, and dynamic querying.

Figure 1 (100) is a block diagram of one preferred system architecture showing one or more computers used by shoppers (101), an online store on the Internet (103), a computer network (102) which is used for communication between the user computers (101) and the online store (103), the sessionization process (109) that converts web server log data (400) from the online store (103) into session data (500), the shopping step finder process (111) that converts session data (500) into micro-conversion data (600), and the visualization process (113) that converts micro-conversion data (600) into visualizations of shopping activities (800) by using a parallel coordinate system.

When a shopper intends to find certain product information and/or purchase one or more products by using the Internet (102), he/she starts a Web browser program (105) in his/her computer (101) and accesses an online store (103) on the Internet which sells the products of interest. Online

stores (103) are implemented with a Web server system (104) which receives Web page requests (106) from shoppers and sends out requested Web pages (107) back to shoppers. For this communication between the shopper's computer (101) and the online store (103), the Web browser program (105) and Web server (104) typically use HTTP (HyperText Transfer Protocol), which is a network protocol defined and implemented for this purpose. A shopper navigates an online store (103), i.e., requests Web pages over the network (102) and browses Web pages which provide information of one or more products sold in the store.

The Web server system (104) in the online store (103) records shoppers' navigation of the online store (103) in its Web server log (400). Namely, the Web server system (104) records all the requests (106) it receives from shoppers in the Web server log (400) along with auxiliary data such as timestamp (401), session ID (402), referrer (403), and links shown in the requested Web page (405) for each request (106). The details of these auxiliary data will be described later in Figure 4 (400). In a preferred embodiment a sessionization process (109) receives a group of requests (106) stored in the Web server log (400) as input and extracts sessions (500) from the requests. In this embodiment, the sessionization process (109) comprises sorting the Web server log (400) by session ID (402) and then by timestamp (401). Other embodiments are envisioned.

A session is a series of Web page requests (106) by one shopper in a single visit. The shopping step finder process (111) is described in detail in Figure 6A. In general, the shopping step finder process determines if there is a request that matches one of the axes of the parallel coordinate system and what the value is of the independent variable to be plotted on the respective axis. For example, the finder process takes sessions (500) extracted by the sessionization process (109) as

input and identifies products seen in each shopping step of each session (500) in the input. The output of the shopping step finder process (111) is referred to as micro-conversions (600) and, in this example, shows how many products seen in one shopping step are converted into the next shopping step.

5

In a non-limiting example, one preferred group of shopping steps includes product impression (604), clickthrough (605), basket placement (606), and purchase (607) and will be described in detail in Figure 6 (600).

For a more detailed description of micro-conversion, refer to U.S. patent application number 09/238,861 to J. Lee, L. Morgenstern, M. Podlaseck, E. Schonberg, and D. Wood, entitled "A System and Method for Collecting and Analyzing Information About Content Requested in a Network (World Wide Web) Environment", filed on January 27, 1999 and U.S. patent application number 09/238,348 to J. Lee, L. Morgenstern, M. Podlaseck, E. Schonberg, and D. Wood, entitled "Aggregating and Analyzing Information about Content Requested in an E-Commerce Web Environment to Determine Conversion Rates", file on January 27, 1999.

Finally, the visualization process (113) takes as input the micro-conversions (600) generated by the previous process (111) and generates as output one or more visualizations of shopping activities (800) in the online store by using a parallel coordinate system. The visualization process (113) plots the independent variable value on the respective axis for each request that has a matching axis in the parallel coordinate system. In a preferred embodiment, the visualization process also generates the polygonal line (using known techniques) to connect the plotted points

on each of the respective axes. In alternative embodiments, the visualization process (113) plots lines of different types (e.g. different line thickness, color, type, etc.) to further sub-categorize the plotted features. For example, sessions started by different customer classes can be designated by different color lines. Sessions also can be categorized by other criteria such as the type of the visited online store, and the type and version of the used browser program and/or operating system.

See U.S. Patent Application number xxx entitled **SYSTEM AND METHOD FOR VISUALLY ANALYZING CLICKSTREAM DATA WITH A PARALLEL COORDINATE SYSTEM**, filed on the same day as this application, to Juhnyoung Lee et al., which is herein incorporated by reference in its entirety.

Figure 2 (200) is a flow chart of a preferred business process. The current invention which provides visualizations of shoppers' activities (800) in an online store (103) can be used by merchants of the online store to understand the effectiveness of their store and/or identify one or more bottlenecks of their store. The merchants can use this information to adjust their Web design and/or merchandising and marketing strategies to improve the performance of their store. The first step (202) of this business process starts with the Web server system (104) of the online store (103) recording all requests (106) from shoppers in its Web server log (400). Then (203), data stored in the Web server log (400) is periodically, say, daily, and/or on demand, retrieved for analysis. Next (204, 205 and 206, respectively), as explained earlier, the Web server log data (400) is processed by the sessionization process (109), the shopping step finder process (111),

and the visualization process (113) to generate one or more visualizations of shoppers' activities (800) in the online store (103).

In some preferred embodiments, the generated visualizations (800) typically come with various filters which business analysts can interactively use to select and/or deselect one or more groups of data for their needs. For example, the business analysts can identify where the store loses customers and how many customers are lost, understand the effectiveness of different merchandising tactics, and understand and compare the shopping behavior of different groups of shoppers, e.g., shoppers from different referrers (Web sites which lead the shopper to this online store), shoppers from different ISPs (Internet Service Providers), and/or shoppers who use different methods to find product information in the store (e.g., hierarchical browsing, keyword search, parametric search, and/or recommendations). The use of filters with the visualizations (800) will be described in detail in Figures 9 (900), 6 (600), and 6A (650).

Next (207), by examining the generated visualizations of shoppers' activities (800) in the online store (103), business analysts understand the effectiveness of their store and/or identify one or more problems with their store such as a broken link to promoted products, or a lengthy and cumbersome checkout process. Business analysts make recommendations for store improvement based their findings.

By using the analysis results from the business analysts, marketers, merchandisers, and Web designers of the online store (103) develop and/or update their strategies for Web design, marketing and merchandising (208), and generate appropriate recommendations for the

new/updated strategies. Finally (209), the recommendations from the Web designers, marketers and merchandisers are passed on to the Web programmers and administrators of the online store (103), who implement the recommended changes to the Web appearance and/or navigation paths of the online store (103) and other Web presentation of the store's marketing and merchandising efforts such as advertisement banners and other links in portal sites such as Yahoo!, AltaVista, Lycos, and Excite@Home.

Figure 3 (300) is an example of a prior art online store structure showing all the Web pages shoppers of the store (103) can request. They include Home page (301), Product category A (302), Product category B (303), Product P1 (304) which is promoted from the Home page (301), Search (305) and Search Result page (311), Product A1 (306) and A2 (307) which belong to product category A (302), Product B1 (308), B2 (309) and B3 (310) which belong to product category B (303), Shopping basket page (312), Purchase page (313), and Thank you page (315). An arrowhead line between two boxes in the figure mean that there a link from the source page to the destination page so that a shopper can directly access the destination page from the source page. That is, a Web page contains hyperlinks in it to all the destination pages which are accessible from this page. For example, Product P1 page (304) is directly accessible from Home page (301), but not from Product A1 page (306). Also, Home page (303) contains hyperlinks to four different Web pages, Product category A (302), Product category B (303), Product P1 (304), and Search page (305). Product P1 page (304) contains only one hyperlink to Basket page (312). Note that Search Result page (311) is a dynamic page whose content changes based on the keyword(s) used in the Search page (305). The Search Result page (311) can contain zero or more hyperlinks to product pages in addition to a hyperlink to Basket page (312).

Figure 4 (400) is an example of a Web server log of an online store showing a number of requests (106) to an online store (103) whose structure was shown in Figure 3 (300). Each request comprises several attributes including timestamp (401), session ID (402), referrer (403), requested page (404), and links (405). Timestamp (401) is the system-generated time when the request was made to the Web server system (104). In this log data (400), requests are sorted by timestamp in an ascending order. Session ID (402) is a unique identification number for different sessions. As explained earlier, a session is a series of Web page requested by a shopper in a single visit. In this log data (400), there are three sessions whose ids are S1, S2 and S3. Requests from these three sessions are interleaved in the log by their timestamp. A referrer (403) is the Web page the shopper was on when he/she makes this request. For example, the referrer of the first request of session S1 (made at T1) is R1, an external Web page, while the referrer of the second request of S1 (made at T3) is Home page which was the current page (404) of the previous request. Links (405) are hyperlinks contained in the current page (404). For example, at T1, the links (405) of the current page (404) , i.e., Home page (301) are Product category A (302), Product category B (303), Product P1 (304), and Search page (305). At T9, the links (405) of the current page (404), i.e., Search Result page (311) are hyperlinks to three product pages, i.e., A2 (307), B1 (308), and B2 (309) which are dynamically generated by the search function of the online store. The use of these links (405) in the server log (400) is novel.

Figure 5 (500) is an example of sessions in an online store showing three sessions extracted from Web server log (400). The sessionization process (109) extracts sessions from requests (106) in Web server log (400) by using timestamp (401) and session ID (402). In this example, the process

(109) extracts three sessions S1 (501), S2 (502) and S3 (503). The structure of each session is the same as that of Web server log (400) having timestamp (501), session ID (505), Referrer (506), current page (507) and links (508). The requests in each session is the same as those in Web server log (400) except they are now separated in different session tables, one table for each session.

Figure 6 (600) is an example of micro-conversions in an online store showing conversions of different sessions (i.e., shoppers) (500) among different shopping steps. The shopping step finder process (111) identifies products seen in each shopping step in consideration. One preferred group of shopping steps in online stores comprises the following four steps: product impression (the view of hyperlinks to a Web page presenting a product) (604), clickthrough (the click on the hyperlink and view the Web page of the product) (605), basket placement (the placement of the item in the shopping basket) (606), and purchase (the purchase of the item, i.e., completion of the transaction) (607). Micro-conversion means a shopper moving to a next shopping step for a product. In this example, the shopping step finder (111) generates one micro-conversion table for each of sessions S1 (601), S2 (602), and S3 (603). For example, S1 (601) has product impressions at five times, T1, T3, T10, T12, and T14. At T1, S1 sees an impression of P1. At T3, S1 sees impressions of A1 and A2. At T10, S1 sees an impression of P1. At T12, S1 sees impressions of B1, B2, and B3. At T14, again S1 sees impressions of B1, B2, and B3. Among these product impressions, only three are converted to clickthroughs, i.e., A1 at T3, B2 at T12, and B3 at T14. Then, two clickthroughs out of these three are converted to basket placements, A1 and B3. Finally, both basket placements, i.e., A1 and B3, are converted into purchases. The

shopping step finder process (111) identifies micro-conversions of sessions S2 and S3 in a similar way, and generates the micro-conversion tables (600(b) and 600(c)).

Figure 6A (650) is a flow chart of a shopping step finder process (111). The process starts with two types of input: a set of requests in a session denoted as $\{R\}$, and a set of (shopping) steps, denoted as $\{S\}$, that the process will identify in the requests belongs to $\{R\}$. The elements of $\{S\}$, i.e., (shopping) steps, provide criteria for deciding if a request belongs to each of the steps. For example, a request in a session belongs to the step of product impressions (604) if the Web page of the request (507) contains one or more hyperlinks (508) to a product page. Also, a request belongs to the step of clickthroughs (605) if the current page of the request (507) is a product page. A request belongs to the step of basket placement (606) if the current page of the request (507) is a basket page. Finally, a request belongs to the step of purchase (607) if the current page of the request (507) is a purchase page.

The shopping step finder process (111 and 650) checks which request belongs to what step by using the decision criteria of each step (655). When the process finds a request belonging to a step, it tags the request as an element of the step (656). After going through this checking for every element in $\{R\}$, i.e., every request in the given session, for every step in $\{S\}$, the process outputs another set $\{C\}$ whose elements C_i represent micro-conversions comprising one or more requests tagged as a particular step, S_i , of $\{S\}$ (659). The shopping step finder process (111 and 650) is executed for every session (500) to identify shopping steps of the sessions that will be visualized by a parallel coordinate system (113).

Figure 7 (700) is an example of a prior art parallel coordinate system which was developed for displaying multivariate data sets to identify the relationship among the variables in the set. A parallel coordinate system (700) comprises a series of parallel lines (701, 702, 703, and 704) that are placed equidistantly. Each parallel axis is assigned a specific dependent variable (e.g., X (701), Y (702), Z (703) and W (704)) and dependent variable values (705) are plotted along the respective axis. The independent variable is represented by polygonal lines (706) which connect the corresponding dependent variable values (also referred to as data points) (705) and which illustrate a relationship between an independent variable and the dependent variables appearing on each axis.

Figure 8 (800) is an example of micro-conversion visualization with a parallel coordinate system. In this figure, each polygonal line represents a single session (501, 502 or 503) and its progression in an online store (103). The first parallel axis (801) represents a session categorizer, in this example, the initial referrer (506) of each session. There are two referrers in the figure, i.e., A (806) and B (807). The next four parallel axes represent shopping steps in an online store (103), i.e., product impression (802), clickthrough (803), basket placement (804), and purchase (805). For data points in the shopping step axes, timestamps of sessions are used. An advantage of using timestamps for data points is that because they are unique to individual sessions, no two sessions share the same data points in these axes. Using data points unique to individual sessions prevents the problem of overlaying lines between two parallel axes, which is serious in parallel coordinate visualizations, because it sometimes obscures the accuracy of visualizations. One disadvantage of using timestamps for data points is that they do not carry any sense of volume. Namely, the existence of a data point in a shopping step axis does not tell how many products were viewed,

placed in basket, or purchased. Rather, it merely says that one or more products were viewed, placed or purchased.

In this figure, it is important to notice that some polygonal lines (808, 809, 810, 812, and 813) stop before they reach the last shopping step, i.e., purchase (805). The parallel coordinate system in this invention does not give a session a data point value for the next shopping step and beyond if the session does not convert to the next step. Hence, each polygonal line stops at the last step the session reached, which indicates the point the session leaves the store. This figure (800) shows that the number of lines connecting two adjacent parallel axes decreases, as polygonal lines go from left to right up to the purchase line (805). Dropouts of polygonal lines visualize where the store loses its customers.

Another note about this figure is that each polygonal line representing a session can be associated with one or more hyperlinks. Clicking on these hyperlinks brings up one or more Web pages providing more information about the associated session such as information about the session owner and products pursued in the session. This feature is referred to as hyperlink association.

Figure 9 (900) is an example of micro-conversion visualization with filters. The parallel coordinate system in this invention can have zero or more filters which help users interactively select and/or de-select one or more groups of sessions and customers in the visualizations. A filter can be also related with the first parallel axis (901) which is a session categorizer. In this figure, there are two filters, i.e., referrer (910) and product category (920). The referrer filter (910) has three values, referrers A, B and C. In this example, referrers A and B are selected, but C is not.

This referrer filter is also represented with the first parallel axis and two values, A (906) and B (907), are displayed. All the sessions visualized in this parallel coordinate system come from either A or B Web site. Real-world examples of referrers of interest may include portal sites such as Yahoo!, AltaVista, Lycos, and Excite@Home. The product category filter has four values, and two values among them, Men's and Women's are selected. Namely, all the sessions visualized in this figure see products from either the Men's or Women's category. In order to distinguish sessions which see Men's products from those which see Women's products, the parallel coordinate system may use different colors to display polygonal lines. For example, it uses blue to display polygonal lines of sessions which see Men's products, while it uses red for sessions which see Women's products.

Figure 10 (1000) is an example of a session browsing more than one online store. When a shopper tries to find one or more products of interest either on the Internet or in the physical world, he/she often visits more than one store to browse and compare products sold in different stores. By aggregating Web server logs (400) from one or more online stores (103) with shoppers' session IDs commonly shared across multiple online stores, an extended sessionization process (109) can construct sessions browsing multiple online stores. In this figure (1000), a session denoted as S3 (1002), visits three stores, first, ST1, then, ST2, and finally, ST3 (1003). With the exception of the store column (1003), the session table (1000) contains the same attributes for sessions, i.e., timestamp (1001), session ID (1002), referrer (1004), current page (1005) and links (1006).

Figure 11 (1100) is an example of micro-conversions of a session browsing more than one online store. From the extended session table shown in Figure 10 (1000), an extended shopping step

finder process (111) can generate a micro-conversion table for a session visiting multiple online stores. This figure (1100) shows the product impressions (1101), clickthroughs (1102), basket placement (1103), and purchases (1104) of the session, S3, shown in Figure 10 (1000). Unlike the micro-conversion table for one online store (600), this table has store name, such as ST1, ST2, or ST3, in each data entry as well as timestamp and product name.

Figure 12 (1200) is an example of visualizing sessions visiting more than one online store. This figure visualizes the progression of two sessions, Session X (1210) and Session Y (1211), in four stores, Store 1 (1206), Store 2 (1207), Store 3 (1208), and Store 4 (1209). The first parallel axis (1201) in this parallel coordinate system uses store as the session categorizer. The next four axes (1202, 1203, 1204 and 1205) are the four shopping steps, i.e., product impressions, clickthroughs, basket placement and purchase, as before. Note that the data points in the shopping step axes are timestamps of the corresponding activities. The visualization illustrates that Session X (1210) first visited Store 1 (1206) but left the store at the clickthrough step (1212). Then it (1210) visited Store 2 (1207) and again left the store at the clickthrough step (1213). Then Session X (1210) visited Store 3 (1208) and purchased from the store (1215). On the other hand, Session Y (1211) first visited Store 2 and left the store at the basket placement step (1214). Then it (1211) visited Store 4 (1209) but left the store at the clickthrough step (1216). Finally, Session Y (1211) visited Store 1 (1206) and purchased from the store (1217).

CLAIMS

We claim:

5 1. A method of doing business on a network comprising:

providing a user with a means to extract one or more sessions from one or more Web server logs
of one or more Web server systems of one or more online stores;

10 providing the user with a means to derive one or more micro-conversions from one or more
sessions from one or more online stores;

providing the user with a means to visualize clickstream data from one or more
micro-conversions;

15 providing the user with a means to interactively request one or more variations of the one or more
clickstream data visualizations;

20 providing the user with a means to interactively generate and view one or more variations of the
one or more clickstream data visualizations upon the user's request; and

providing the user with a means to store one or more generated clickstream data visualizations in
one or more computer memories.

2. A method, as in claim 1, where the micro-conversion is a shopper's conversion from one shopping step to another for a particular product or service.

5 3. A method, as in claim 2, where the shopping steps include a product impression that is the view of a hyperlink to a Web page presenting a product or service, a clickthrough that is the click on the hyperlink and view of the Web page of the product or service, a basket placement that is the placement of the item in the shopping basket, and a purchase that is the purchase of the item and the completion of the transaction.

10 4. A method, as in claim 1, where the clickstream data is a collection of micro-conversions of one or more shoppers for one or more products and/or services sold in one or more online stores.

15 5. A method, as in claim 1, where the visualization of clickstream data comprises a traditional parallel coordinate system and one or more extension components including one or more parallel axes of sequential events, one or more dependent variable values of timestamps, one or more dropouts of polygonal lines, one or more filters, one or more categorizers, and one or more hyperlink associations.

20 6. A method, as in claim 5, where the traditional parallel coordinate system is a parallel coordinate system comprising a series of parallel lines that are placed equidistantly, each parallel line being assigned a specific dependent variable and dependent variable values being plotted along the respective axis, and an independent variable that is represented by polygonal lines connecting the

corresponding dependent variable values and illustrating a relationship between an independent variable and the dependent variables appearing on each axis.

7. A method, as in claim 5, where the parallel axes of sequential events is an assignment of a series of sequential events to parallel lines in a parallel coordinate system.

8. A method, as in claim 7, where the sequential events include any one or more of the following: one or more steps of shopping in one or more stores, and one or more product or service development steps.

9. A method, as in claim 5, where the dependent variable values of timestamps is an assignment of timestamp values as data points to a series of sequential events that are assigned to the equal number of parallel axes in a parallel coordinate system.

10. A method, as in claim 5, where the dropout of a polygonal line is the disappearance of a polygonal line before the line reaches the last parallel axis in a parallel coordinate system with the parallel axes of sequential events.

11. A method, as in claim 5, where the filter is a means to select and/or de-select one or more groups of polygonal lines viewed in a parallel coordinate system.

12. A method, as in claim 5, where the categorizer is a parallel axis in a parallel coordinate system whose purpose is to categorize polygonal lines in the system.

13. A method, as in claim 12, where the categorizer includes one or more of the following: the
referrer Web sites of sessions, the ISPs (Internet Service Providers) of sessions, the lengths of
sessions, the methods used to find product information by sessions, the geographic regions where
sessions come from, the ages, sex, education levels, and income levels of the owners of sessions,
the sales history of the owners of sessions, the Web page patterns accessed by sessions or by the
owners of sessions, either or not ordered by session, or by time.

14. A method, as in claim 5, where the hyperlink association is the association of one or more
hyperlinks with the polygonal line representing a session, clicking on the polygonal line opens a
Web page providing detail information of the session.

15. A method, as in claim 1, where the user can identify where the online store loses customers,
and/or how many customers are lost, by looking at the dropouts of one or more polygonal lines in
the clickstream data visualizations.

16. A method, as in claim 1, where the user can view one or more variations of the clickstream
data visualization by selecting and/or deselecting one or more groups of sessions in the
visualization by using one or more filters.

17. A method, as in claim 1, where the user can view one or more clickstream data visualizations
for sessions of different shoppers categorized by one or more values of the categorizer axis.

18. A method, as in claim 1, where the user can view one or more Web pages providing detail information of one or more sessions by using one or more hyperlink association, i.e., by clicking on one or more polygonal lines representing one or more sessions.

19. A method, as in claim 1, where the user can store one or more snapshots of the clickstream data visualization of the online store, and later compare the stored snapshots of the visualization to understand the changes in the performance of the online store.

20. A method, as in claim 1, where the user develops and/or updates the strategies for the Web design, marketing and merchandising based on the findings from the clickstream data visualizations.

21. A method, as in claim 1, where the user adjusts and/or updates the appearance of the Web design or navigation paths of the online store and/or other Web presentations of the store's marketing and merchandising efforts including advertisement banners and product or services layouts to reflect the developed/updated strategies based on the analysis results.

BUSINESS METHOD FOR VISUALLY ANALYZING CLICKSTREAM DATA WITH A PARALLEL COORDINATE SYSTEM

ABSTRACT

An interactive parallel coordinate system is used in a business method to provide users with greater abilities to interpret and explore the clickstream data of online stores on the Web. The interactive parallel coordinate system uses the multiple axes of the system to represent steps of a process with respect to an independent variable. Specifically, the process steps are sequential steps which sessions take during their navigation of a Web site.

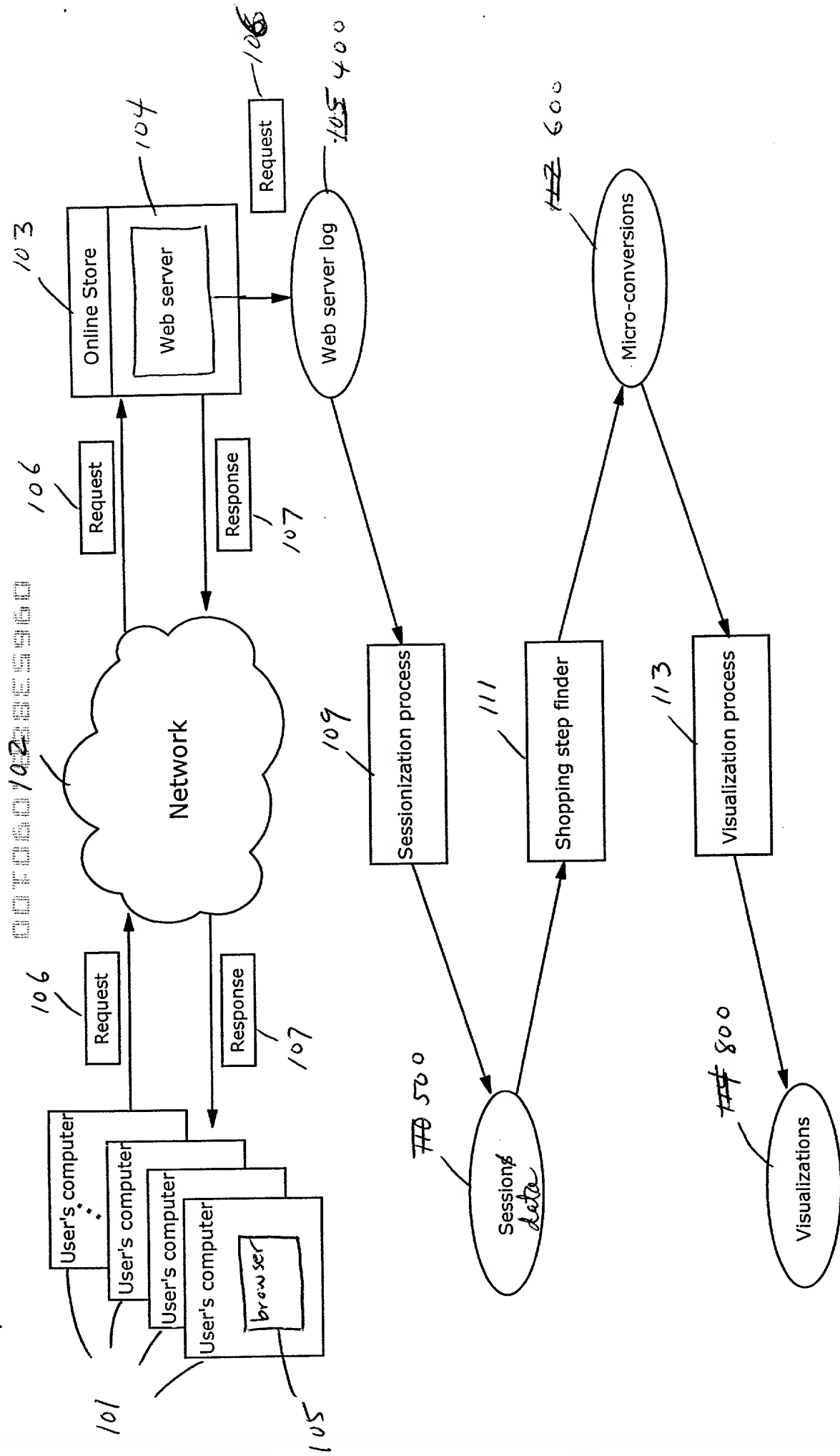


Figure 1: A block diagram of one preferred system architecture

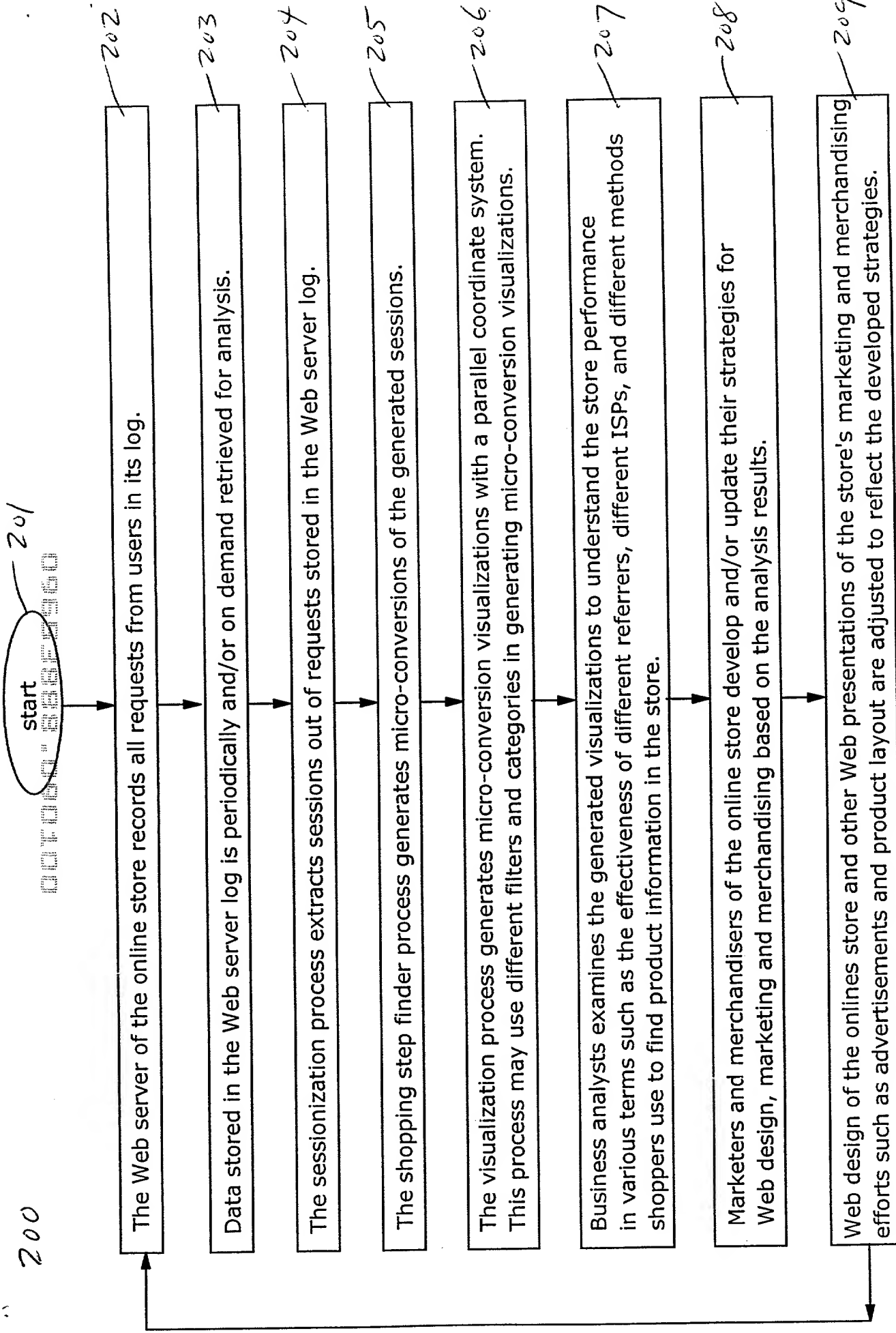
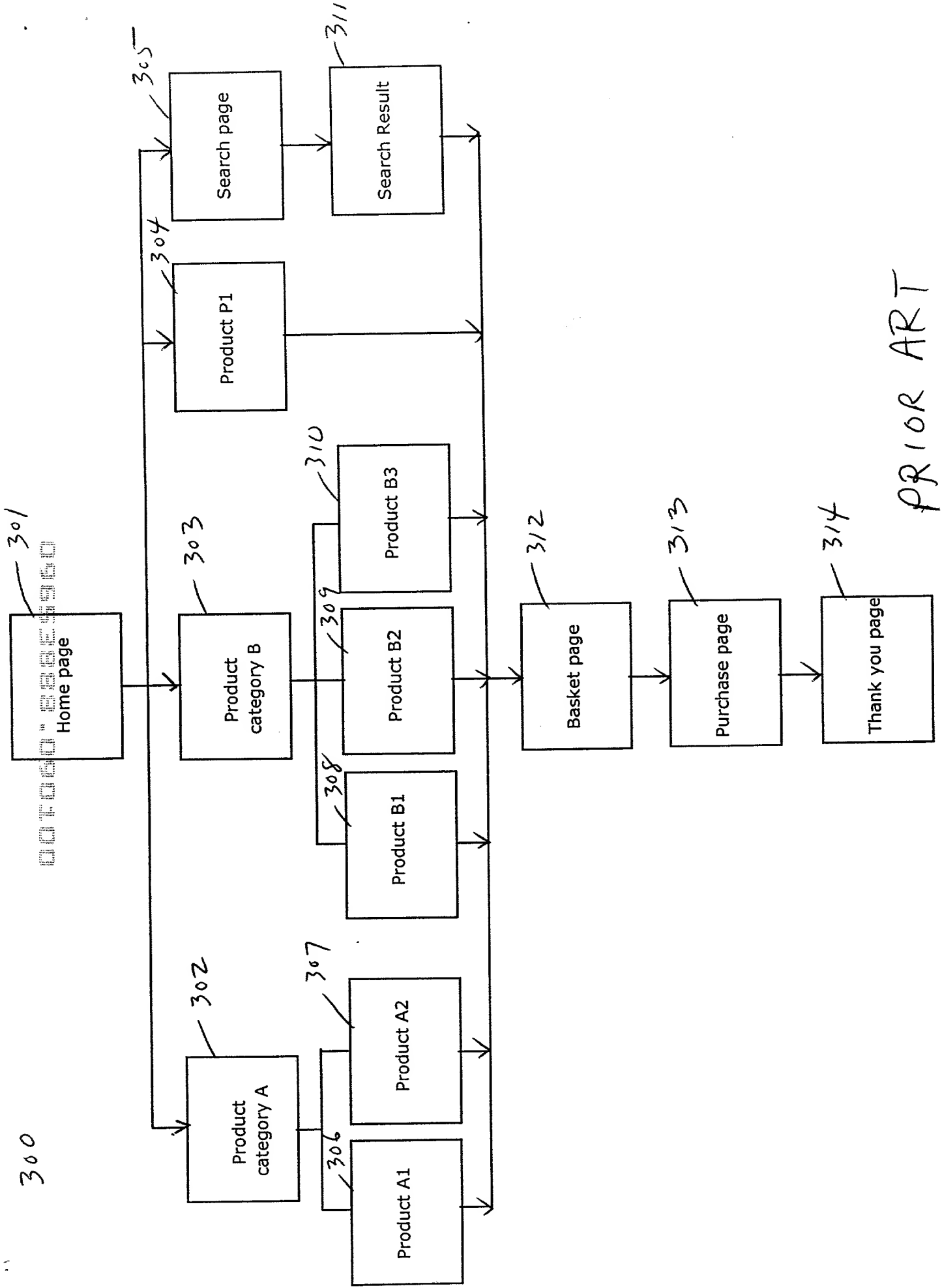


Figure 2: A flow chart of a preferred business process



PRIOR ART

Figure 3: An example of an online store structure

500

501

502

503

504

505

506

507

508

509

510

511

512

513

514

515

516

517

518

519

520

521

522

523

524

525

526

527

528

529

530

531

532

533

534

535

536

537

538

539

540

541

542

543

544

545

546

547

548

549

550

551

552

553

554

555

556

557

558

559

560

561

562

563

564

565

566

567

568

569

570

571

572

573

574

575

576

577

578

579

580

581

582

583

584

585

586

587

588

589

590

591

592

593

594

595

596

597

598

599

600

601

602

603

604

605

606

607

608

609

610

611

612

613

614

615

616

617

618

619

620

621

622

623

624

625

626

627

628

629

630

631

632

633

634

635

636

637

638

639

640

641

642

643

644

645

646

647

648

649

650

651

652

653

654

655

656

657

658

659

660

661

662

663

664

665

666

667

668

669

670

671

672

673

674

675

676

677

678

679

680

681

682

683

684

685

686

687

688

689

690

691

692

693

694

695

696

697

698

699

700

701

702

703

704

705

706

707

708

709

710

711

712

713

714

715

716

717

718

719

720

721

722

723

724

725

726

727

728

729

730

731

732

733

734

735

736

737

738

739

740

741

742

743

744

745

746

747

748

749

750

751

752

753

754

755

756

757

758

759

760

761

762

763

764

765

766

767

768

769

770

771

772

773

774

775

776

777

778

779

780

781

782

783

784

785

786

787

788

789

790

791

792

793

794

795

796

797

798

799

800

801

802

803

804

805

806

807

808

809

810

811

812

813

814

815

816

817

818

819

820

821

822

823

824

825

826

827

828

829

830

831

832

833

834

835

836

837

838

839

840

841

842

843

844

845

846

847

848

849

850

851

852

853

604 605 606 607

Product impressions	Clickthroughs	Basket placement	Purchase
T1, P1	T6, A1	T8, A1	T20, A1
T3, A1	T13, B2	T19, B3	T20, B3
T3, A2	T16, B3		
T10, P1			
T12, B1			
T12, B2			
T12, B3			
T14, B1			
T14, B2			
T14, B3			

(a) Micro-conversions of session S1 — 601

Product impressions	Clickthroughs	Basket placement	Purchase
T2, P1	T11, B1	T15, B1	T17, B1
T9, A2			
T9, B1			
T9, B2			

(b) Micro-conversions of session S2 — 602

Product impressions	Clickthroughs	Basket placement	Purchase
T4, P1	T7, P1		
T18, P1			
T22, A1			
T22, A2			

(c) Micro-conversions of session S3 — 603

Figure 6: An example of micro-conversions in an online store

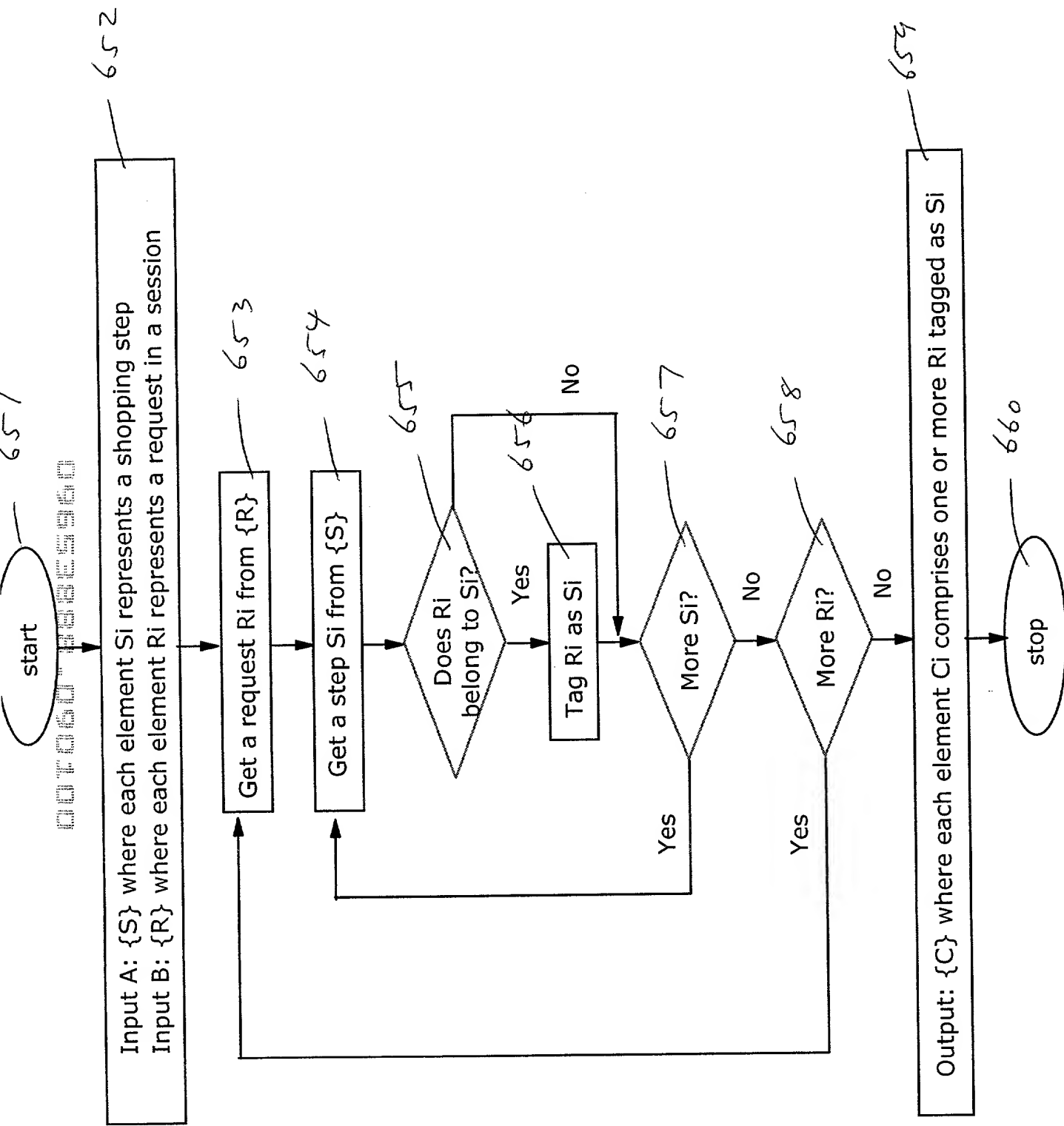
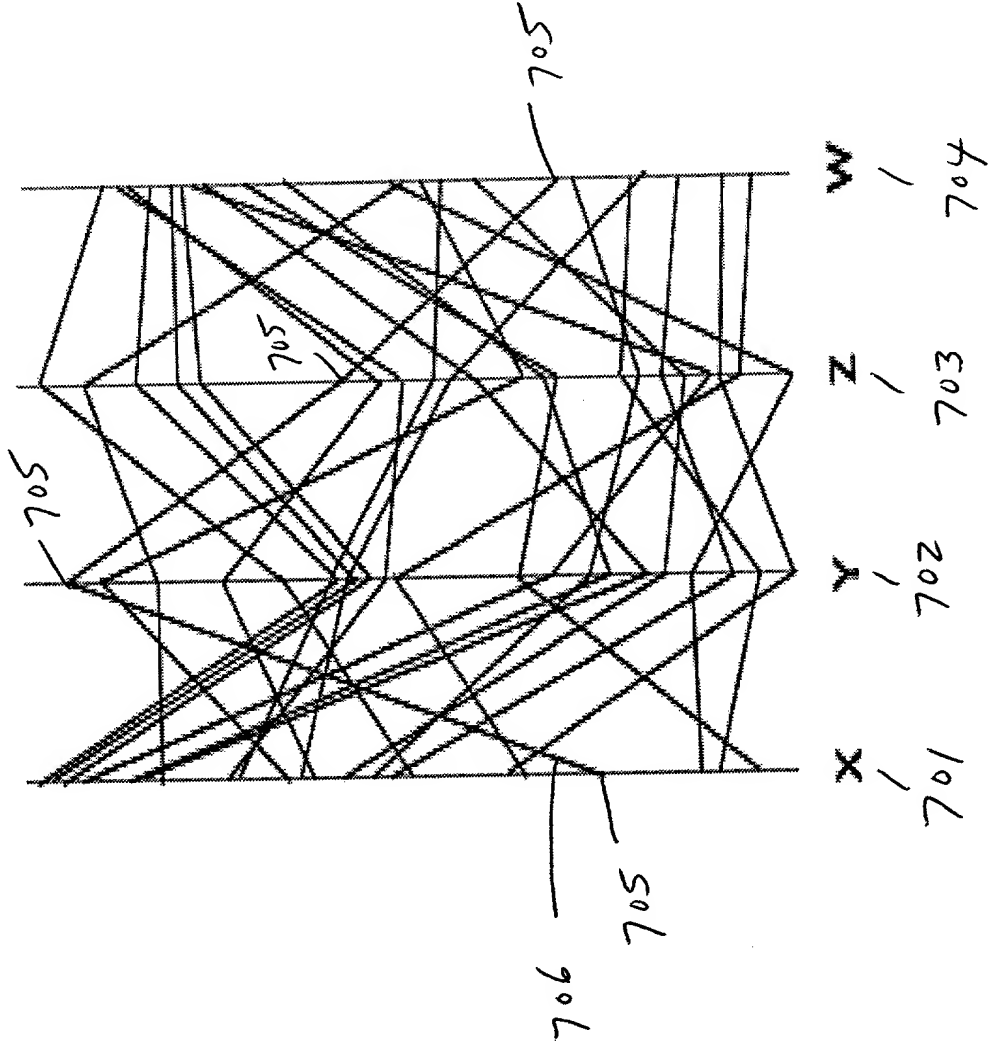


Figure 6A: A flow chart of a shopping step finder process

700



Prior Art

Figure 7: An example of a traditional parallel coordinate system

800

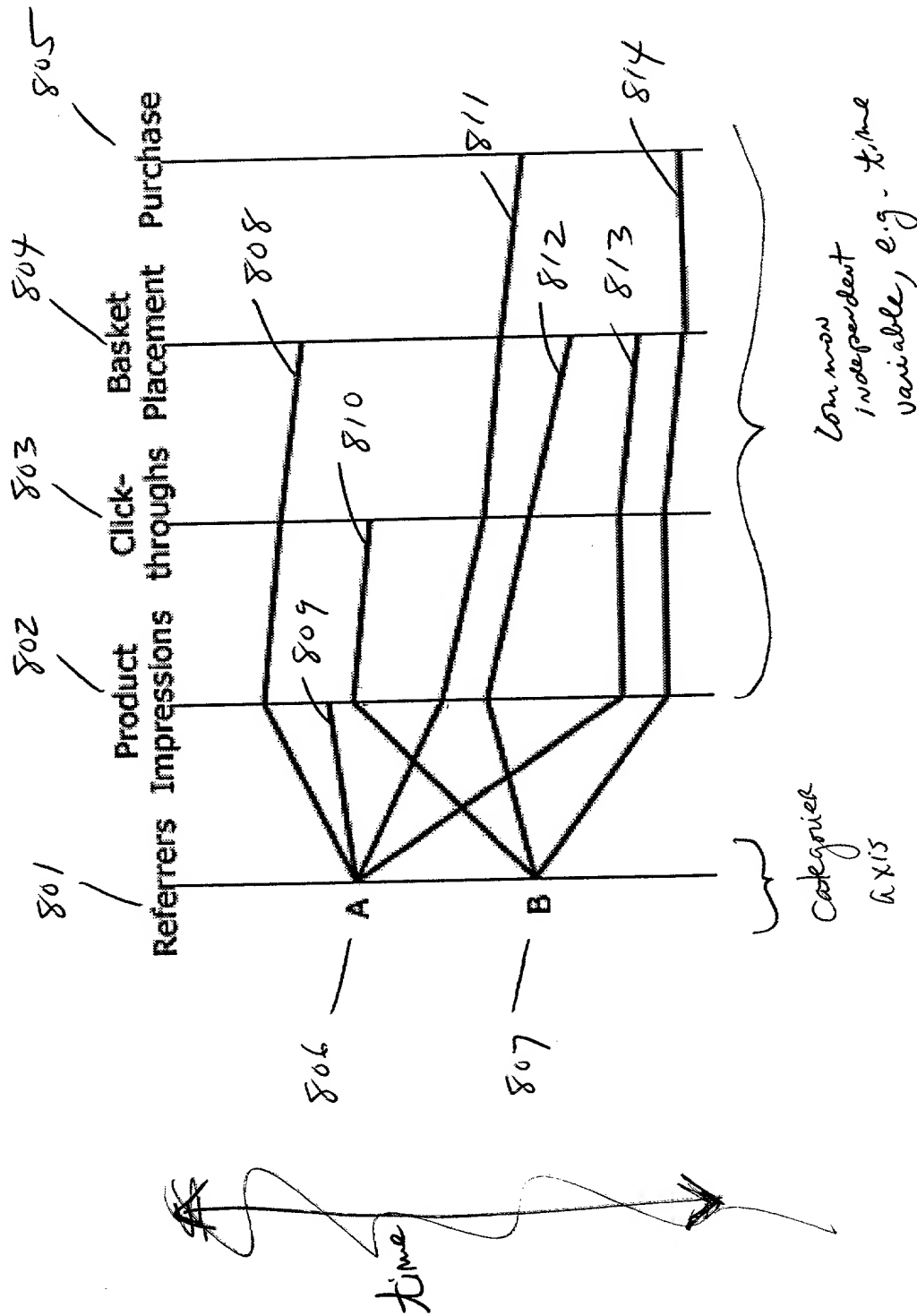


Figure 8: An example of a micro-conversion visualization with a parallel coordinate system

900

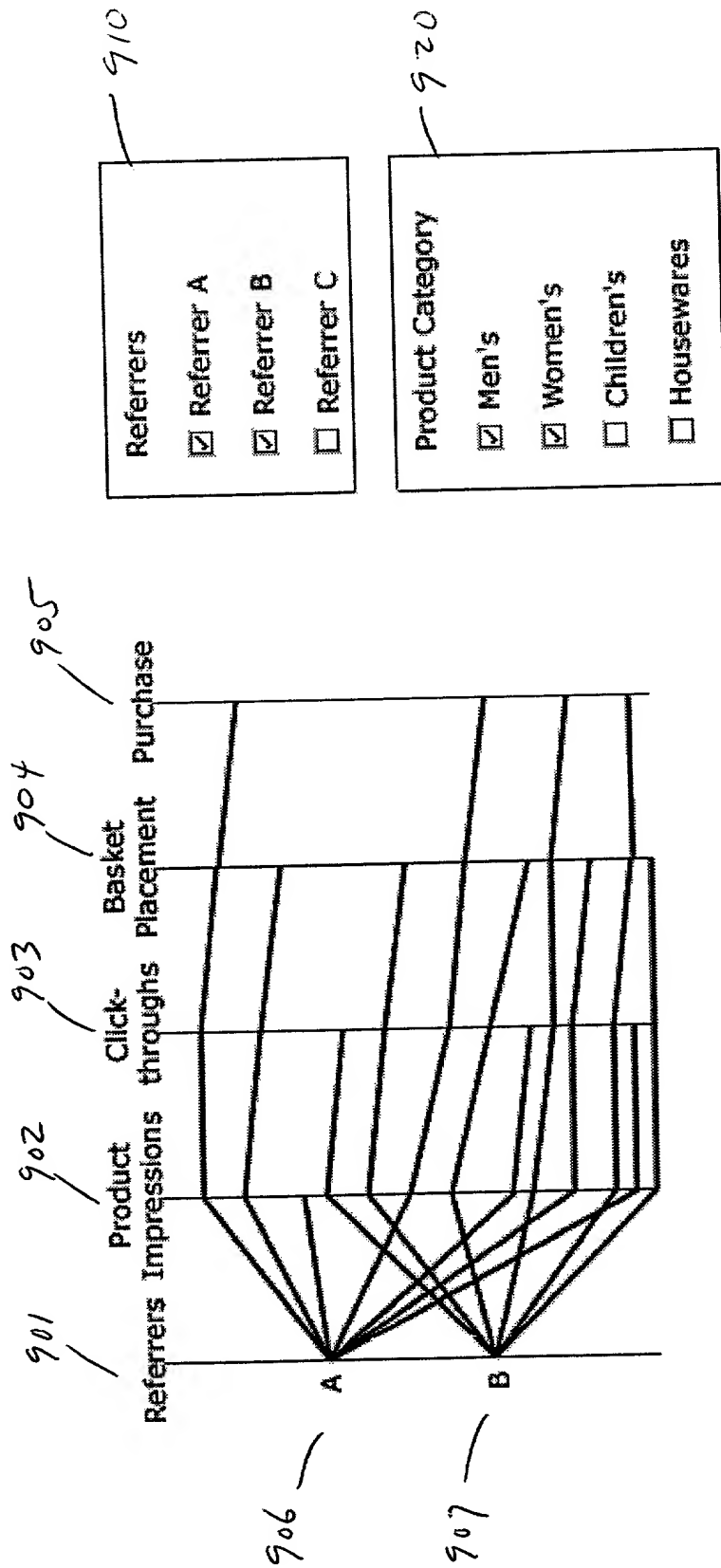


Figure 9: An example of a micro-conversion visualization with filters

001050" 888E5560

1000

1006

1005

1004

1003

1002

1001

Timestamp	Session ID	Store	Referrer	Current Page	Links
T4	S3	ST1	R3	Home	A, B, P1, Search
T7	S3	ST1	Home	P1	Basket
T18	S3	ST1	P1	Home	A, B, P1, Search
T22	S3	ST1	Home	A	A1, A2
T23	S3	ST2	ST1	Home	B, C, D, Search
T24	S3	ST2	Home	Search	
T25	S3	ST2	Search	Search Results	C2, C3, D1
T26	S3	ST2	Search Results	C2	Basket, C
T27	S3	ST2	C2	Search Results	C2, C3, D1
T28	S3	ST2	Search Results	C3	Basket, C
T29	S3	ST3	ST2	Home	A, B, C, D, Search
T30	S3	ST3	Home	Search	
T31	S3	ST3	Search	Search Results	A1, B1, C3, D2
T32	S3	ST3	Search Results	D2	Basket, D
T33	S3	ST3	D2	Basket	Purchase, Home
T34	S3	ST3	Basket	Home	A, B, C, D, Search
T35	S3	ST3	Home	D	D1, D2, D3
T36	S3	ST3	D	D3	Basket, D
T37	S3	ST3	D3	Basket	Purchase, Home
T38	S3	ST3	Basket	Purchase	Thank You
T39	S3	ST3	Purchase	Thank You	Home

Figure 10: An example of a session browsing more than one online stores

1100

DOTCOM" 888E9960

1101

1102

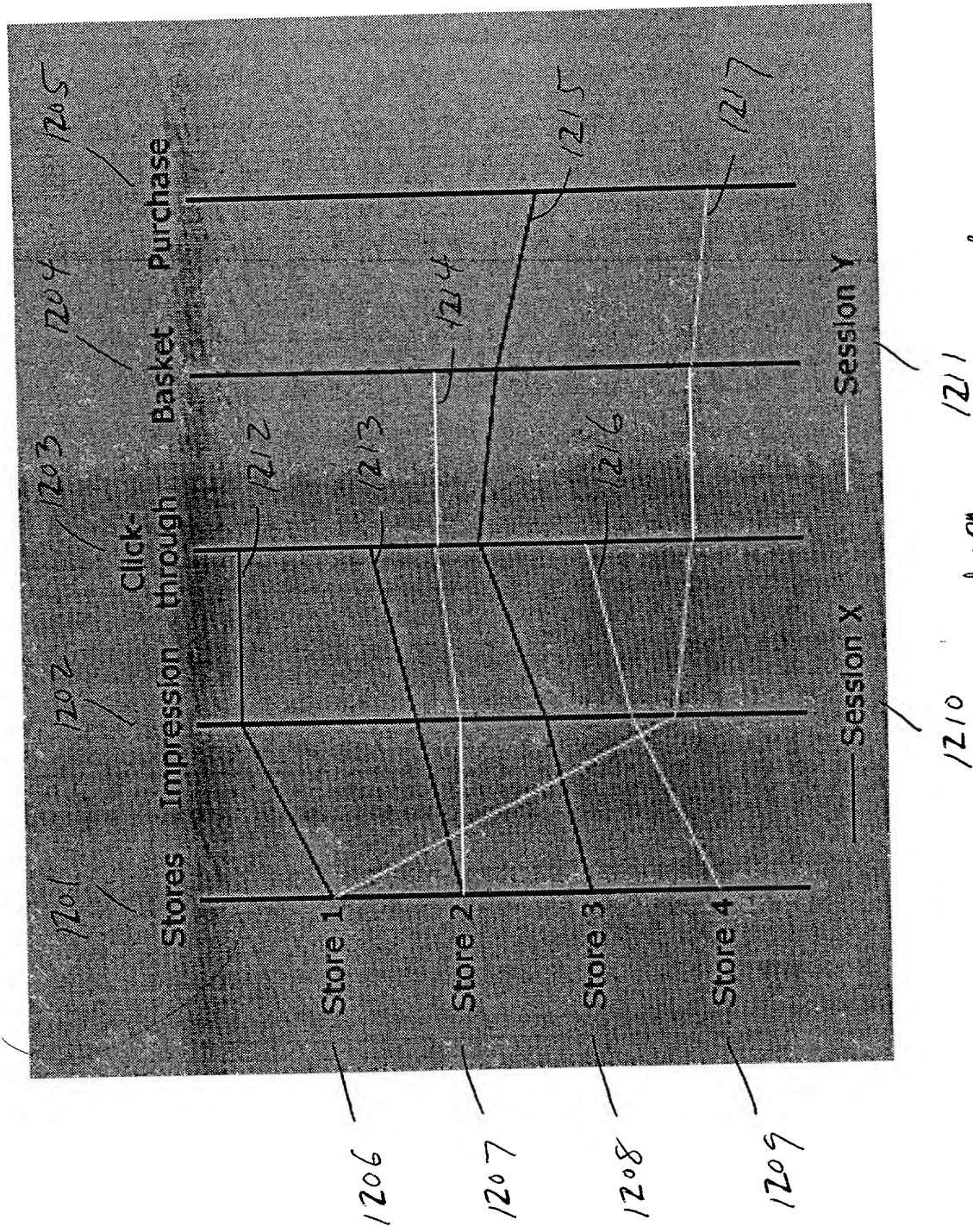
1103

1104

Product impressions	Clickthroughs	Basket placement	Purchase
T4, P1, ST1	T7, P1, ST1	T33, D2, ST3	T38, D2, ST3
T18, P1, ST1	T26, C2, ST2	T37, D3, ST3	T38, D3, ST3
T22, A1, ST1	T28, C3, ST2		
T22, A2, ST1	T32, D2, ST3		
T25, C1, ST2	T36, D3, ST3		
T25, C3, ST2			
T25, D1, ST2			
T27, C2, ST2			
T27, C3, ST2			
T27, D1, ST2			
T31, A1, ST3			
T31, B1, ST3			
T31, C3, ST3			
T31, D2, ST3			
T35, D1, ST3			
T35, D2, ST3			
T35, D3, ST3			

Figure 11: An example of micro-conversions of a session browsing more than one online stores

1200
 001060" 8885960
 Category 5200



Sub-category
 by this session = set of events

Figure 12: An example of visualizing shopping activities in more than one online stores

DECLARATION AND POWER OF ATTORNEY FOR PATENT APPLICATION

As a below named inventor, I hereby declare that:

My residence, post office address and citizenship are as stated below next to my name;

I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought on the invention entitled:

BUSINESS METHOD FOR VISUALLY ANALYZING CLICKSTREAM DATA WITH A PARALLEL COORDINATE SYSTEM

the specification of which (check one)

X is attached hereto.

_____ was filed on _____ as United States Application Number

or PCT International Application Number _____

and was amended on _____ (if applicable)

I hereby state that I have reviewed and understand the contents of the above identified specification, including the claims, as amended by any amendment referred to above.

I acknowledge the duty to disclose information which is material to the patentability of this application in accordance with Title 37, Code of Federal Regulations, Section 1.56.

I hereby claim foreign priority benefits under Title 35, United States Code, §119(a)-(d) or §365(b) of any foreign application(s) for patent or inventor's certificate, or §365(a) of any PCT International application which designated at least one country other than the United States, listed below and have also identified below, by checking the box, any foreign application for patent or inventor's certificate, or PCT International application, having a filing date before that of the application on which priority is claimed:

Prior Foreign Application(s)			Priority Claimed	
_____ (Number)	_____ (Country)	_____ (Day/Month/Year Filed)	_____ Yes	_____ No
_____ (Number)	_____ (Country)	_____ (Day/Month/Year Filed)	_____ Yes	_____ No
_____ (Number)	_____ (Country)	_____ (Day/Month/Year Filed)	_____ Yes	_____ No

I hereby claim the benefit under 35 U.S.C. §119(e) of any United States provisional application(s) listed below.

_____ (Application Number)	_____ (Filing Date)
_____ (Application Number)	_____ (Filing Date)

I hereby claim the benefit under 35 U.S.C. §120 of any United States Application(s), or §365(c) of any PCT International application designating the United States, listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States, or PCT International application in the manner provided by the first paragraph of 35 U.S.C. §112, I acknowledge the duty to disclose information material to the patentability of this application as defined in 37 CFR §1.56 which occurred between the filing date of the prior application and the national or PCT international filing date of this application:

_____ (Application Serial No.)	_____ (Filing Date)	_____ (Status) (patented, pending, abandoned)
_____ (Application Serial No.)	_____ (Filing Date)	_____ (Status) (patented, pending, abandoned)

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code and that willful false statements may jeopardize the validity of the application or any patent issued thereon.

POWER OF ATTORNEY: As a named inventor I hereby appoint the following attorney(s) and/or agent(s) to prosecute this application and transact all business in the Patent and Trademark Office connected therewith (list name and registration number).

Manny W. Schecter (Reg. 31,722), Lauren Bruzzone (35,082), Christopher A. Hughes (Reg. 26,914), Edward A. Pennington (Reg. 32,588), John E. Hoel (Reg. 26,279), Joseph C. Redmond, Jr. (Reg. 18,753), Douglas W. Cameron (Reg. 31,596), Louis P. Herzberg (Reg. 41,500), Stephen C. Kaufman (Reg. 29,551), Daniel P. Morris (Reg. 32,053), Louis J. Percello (Reg. 33,206), David M. Shofi (Reg. 39,835), Robert M. Trepp (Reg. 25,933), Paul J. Otterstedt (Reg. 37,411), Wayne L. Ellenbogen (43,602) and Marian Underweiser (Reg. 46,134)

Send Correspondence to: Louis J. Percello, Intellectual Property Law Dept.IBM Corporation, P.O. Box 218, Yorktown Heights, New York 10598Direct Telephone Calls to: (name and telephone number) Louis J. Percello (914)945-3145Thomas Anthony Cofino**Full name of sole or first inventor**

Inventor's Signature

Date

19 Jean Street, Rye, New York 10580

Residence

USA

Citizenship

same as above
Post Office Address

DECLARATION AND POWER OF ATTORNEY FOR PATENT APPLICATIONStephen H. Gomory**Full name of second joint-inventor, if any**

Inventor's signature

Date

68 Leroy Avenue, Tarrytown, New York 10591

Residence

USA

Citizenship

same as above

Post Office Address

Juhnyoung Lee**Full name of third joint-inventor, if any**

Inventor's signature

Date

211 Treetop Lane, Yorktown Heights, New York 10598

Residence

Republic of Korea

Citizenship

same as above

Post Office Address

Mark Edward Podlaseck**Full name of fourth joint-inventor, if any**

Inventor's Signature

Date

80 Valley Road, New Preston, Connecticut 06777

Residence

USA

Citizenship

same as above

Post Office Address

Full name of fifth joint inventor, if any

Inventor's Signature

Date

Residence

Citizenship

Post Office Address

Full name of sixth joint-inventor, if any

Inventor's signature

Date

Residence

Citizenship

Post Office Address